HOWTO - Basic Analytics using RadioDNS
Last update: 2020-07-01

Introduction

The RadioDNS model means that we hold no content, no transactions pass through any of our servers, and we have only the most vague understanding of the number of DNS lookups being performed due to the aggressive caching nature of DNS records.

That means it is not possible for RadioDNS to provide any analytics information directly.

However, the way that RadioDNS works means that each individual broadcaster and manufacturer can analyse their own system logs to understand usage. That does not require any involvement of other parties, and the information which is captured and analysed remains confidential.

This document suggests ways that RadioDNS activity can be logged and analysed.

Logging of RadioDNS Traffic

The Service and Programme Information standard (TS 102 818) describes how a receiver gets information from the broadcaster, and uses HTTP for transport.

In almost all cases, that means the information is being provided by a webserver like Microsoft IIS, Apache HTTP server, nginx, icecast, wowza or similar. All these servers keep access logs, which record every request, and often in the **Combined Log Format**.

A Combined Log Format entry will typically include:
- The Origination IP Address
- The Date / Time of access
- The **Resource** being requested (e.g. "GET /radiodns/spi/3.1/SI.xml")
- The **User-Agent** string

For example:
```
127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700] "GET /apache_pb.gif HTTP/1.0"
200 2326 "http://www.example.com/start.html" "Mozilla/4.08 [en] (Win98; I
;Nav)"
```
(from https://httpd.apache.org/docs/1.3/logs.html#combined)

By filtering these log entries down, you can see what traffic is originating from RadioDNS devices.

The **Resources** used by RadioDNS devices are:
- Anything beginning "/radiodns/spi/" (Regex: ^\/radiodns\/spi\/.*$)
- Any Logo URLs specified in the SI file
- Any Streaming URLs specified in the SI file

- Any Image URLs triggered by the Visual Slideshow application

You may find it easier to track if you use unique URLs for RadioDNS resources. You could use techniques like:
- Distinct URLs for RadioDNS assets - e.g. "/radiodns/assets/logos/logo_url"
- Distinct parameter addition - e.g. "logo_url?radiodns"
- .htaccess / mod_rewrite redirection - e.g. "/radiodns/logos -> /assets/logos"

You can then Filter the log file to find accesses to the RadioDNS resources over time.

User-Agent

Devices / systems accessing resources should use a User-Agent which uniquely describes the **class** of device. We recommend that manufacturers follow this guidance to create meaningful User-Agent values, and we will offer to maintain a list of which User-Agents correspond to which systems.

For example:
- The iHeartRadio app declares a user agent of "iHeartRadio/"
- The Pandora Radio app on android declared a user agent of "Pandora.+Android"

By analysing the User-Agent values, you can determine who/what is accessing your RadioDNS resources.

The User-Agent header only identifies a **class** of device. It does not uniquely identify each device.

Log Analysis Software

As a typical logfile will have a very large number of lines (potentially a million lines per day), they are best filtered and analysed using specialist log analysis tools. There is a list of such tools, many of which are FOSS (Free Open Source Software) at
https://en.wikipedia.org/wiki/List_of_web_analytics_software

Conditional Responses

You can use rules on your webserver (for example, Apache's mod_rewrite, or IIS's URL Rewrite) to divert requests based on parameters like the User-Agent value, the originating IP address. That makes it possible to serve different versions of resources like the SI file or logos based on **the class of device** which is asking for them, which in turn can make it easier to analyse activity.

Alternatively, you can use a server-side scripting language (like python or node) to generate a resource on the fly, similarly adapting to values provided by the client. This can be more flexible but is generally much slower and more resource intensive, and we don't generally recommend it.

Implications of Caching

Caching will mean that not every request for a resource may be logged on your server. If the device already has a resource cached and the cache time hasn't expired, it won't contact your server again.

You should **always** allow resources to be cached for a reasonable period of time. If you set null or unreasonably low caching times, the device may ignore them or ignore your content.

If you set the caching time for your SI document to 24 hours, you will generally see one hit in your log every 24 hours from each device. Similarly, if you set the caching of your station logos to

90 days, you'll see one hit in your log every 90 days.

Some devices will connect via a proxy server, usually for security reasons. This is often the case for automotive devices. In that case, the proxy server will contact your server and distribute the assets to the devices. You'll see one hit in the log for the proxy server, not multiple hits from each device.

Log analysis of SI files and Logos requests will generally underestimate use due to caching and proxies.

Content not subject to Caching

Some content is not cacheable, so will generate a log entry for every device at every event

Typically this is:
- Streaming events - when a stream ends a log entry is created.
- Visual events - each device will connect to the server when it tunes into a radio station, and disconnect when it tunes out. This can give a very accurate indication of aggregate device activity over time. Additionally, each time a visual is updated, a request is triggered to the server for an image asset which will generate a log event for that image.

Devices are recommended not to try to cache or aggregate these requests, so you should see a request from each device.

However, not all devices support streaming and/or visuals, so the overall response will be lower than the total number of devices supporting SI and logo files.

Log analysis of streaming audio and visuals will generally give an accurate picture of how many devices are active over time, but will underestimate the number of devices in total.

Analytics Developments

The RadioDNS Technology Group is looking at how more accurate analytics information can be provided to broadcasters, whilst staying within constraints imposed by privacy laws and similar regulations.